Labelling and Family Resemblance in the discrimination of polymorphous categories by pigeons

Elizabeth Nicholls, Catriona M. E. Ryan,

Catherine M. L. Bryant, Stephen E. G. Lea

University of Exeter
School of Psychology
Washington Singer Laboratories
Perry Road
Exeter
EX4 4QG

Email: s.e.g.lea@exeter.ac.uk
Tel: +44 1392 264612
Fax: +44 1392 264623

Animal Cognition MS AC-09-0161, Accepted for publication subject to revision

This draft last saved 1 July 2010 by SEGL

Abstract

Two experiments examined whether pigeons discriminate polymorphous categories on the basis of a single highly predictive feature or overall similarity. In the first experiment, pigeons were trained to discriminate between categories of photographs of complex real objects. Within these pictures, single features had been manipulated to produce a highly salient texture cue. Either the picture or the texture provided a reliable cue for discrimination during training, but in probe tests, the picture and texture cues were put into conflict. Some pigeons showed a significant tendency to discriminate on the basis of the picture cue (overall similarity or family resemblance), whereas others appeared to rely on the manipulated texture cue. The second experiment used artificial polymorphous categories in which one dimension of the stimulus provided a completely reliable cue to category membership, whereas three other dimensions provided cues that were individually unreliable but collectively provided a completely reliable basis for discrimination. Most pigeons came under the control of the reliable cue rather than the unreliable cues. A minority, however, came under the control of single dimensions from the unreliable set. We conclude that cue salience can be more important than cue reliability in determining what features will control behaviour when multiple cues are available.

Key words: Pigeon, Category discrimination, Feature learning, Family resemblance, Labelling, Salience

General Introduction

The ability to categorise natural objects is expected to be widespread across the animal kingdom, since it permits a reduction in the amount of information an animal must acquire about its environment in order to respond adaptively (Makino and Jitsumori 2007). Since the pioneering experiments of Herrnstein and Loveland (1964) birds have been shown to be capable of discriminating a wide range of categories, with category members varying from simple artificially constructed stimuli (e.g. Lea and Harrison 1978) to photographs of complex natural-language categories such as trees (Herrnstein et al. 1976; for a review see Huber 2001).

A typical categorisation experiment involves pigeons learning to associate one stimulus set or category with a food reward, when there is no single feature that is necessary or sufficient for category membership. However it is not clear what pigeons learn to associate with the food reward. It could be that the pigeon has elaborated a concept corresponding to the category. Typically however it is assumed that successful category discrimination does not necessarily imply that pigeons have a concept corresponding to the experimenter-defined category (despite the terminology of early experiments e.g. Herrnstein et al. 1976), or in the case of natural photographs, the ability to generalise from images to the corresponding real objects (Bovet and Vauclair 2000). But if pigeons are not using concepts to categorize complex stimuli, how might they be doing it? There are two well studied possibilities. It could be that they learn how to respond to one or more specific exemplars, with other stimuli being categorised by a process of generalisation. Alternatively, it may be that they learn to respond to a collectivity of features that are between them good enough predictors of

category membership, even if none of them is individually necessary or sufficient. These possibilities are usually referred to respectively as exemplar learning and learning by family resemblance or overall similarity.

It is often useful to compare the cognitive abilities of non-human animals and human infants, because in both cases subjects have to complete tasks without the elaborate language competence that facilitates so many cognitive performances in adult humans. The present experiment draws on two results that have been established in the field of infant categorization, in order to pose questions about categorization in pigeons.

Both phenomena involve the use of single, highly predictive features within sets of multidimensional stimuli. However, they are to some extent contradictory. On the one hand, under conditions where adults and older children typically categorise items according to a single stimulus dimension, infants have been found to group items according to overall similarity or family resemblance across numerous stimulus dimensions (Smith and Kemler 1977; Smith 1981). On the other hand, the provision of a verbal label which is invariantly associated with the members of one category has been found to facilitate children's category acquisition (Waxman and Markow, 1995), even among infants too young to have functional speech (e.g. Balaban and Waxman, 1997); so in contrast to the control by family resemblance seen in spontaneous categorization, the presence of a single reliable feature is found to facilitate category acquisition.

The first of these findings, the tendency of younger children to categorize on the basis of overall similarity, has supported a general arguments that such categorization must require simpler cognitive mechanisms than using a single stimulus dimension (e.g. Ashby et al. 1998; Pothos 2005). Unidimensional discrimination is assumed to require the ability to verbalise a rule, something that a pre-verbal infant cannot do. In accordance with this assumption, Couchman et al. (2010) found that, when they used stimulus sets of a kind introduced by Kemler Nelson (1984), which can be discriminated either on the basis of a single dimension or on a family resemblance across three other dimensions, human participants categorised them unidimensionally whereas two rhesus monkey subjects categorized them mainly by family resemblance.

However, it is not the case that non-human animals always categorize complex stimuli by overall similarity rather than unidimensionally, or that unidimensional categorization implies more complex cognitive processing (e.g. Lea and Wills 2008; Wills et al. 2009). Although it is certainly possible for pigeons to use multiple stimulus dimensions in making complex discriminations (e.g. Blough 1972; Herbranson et al. 1999; Kirkpatrick-Steger and Wasserman 1996; Soto and Wasserman 2010; von Fersen and Lea 1990), this is not necessarily what they do most easily or spontaneously. Rather than categorizing in terms of overall similarity, which would involve all available features, birds in categorization experiments with complex stimuli often come under the control of just one or two features (e.g. Lea et al. 1993, 2006); and it would be absurd to suppose that pigeons can verbalise rules (cf. Lea and Wills, 2008). Furthermore, the fact that infant categorization proceeds more easily in the presence of a verbal label – which is really nothing but a single highly valid

stimulus dimension (Plunkett et al., 2008) – suggests that discriminating a single dimension may be a cognitively simpler task than discriminating overall similarity.

The aim of the present study is to explore whether, under conditions where both strategies would be equally effective, pigeons would solve categorization tasks by using a single highly predictive feature, or by using the overall similarity across a group of features of more modest individual predictiveness. However, previous results have shown that when pigeons do make use of single features, they do not always select the most predictive feature: a less valid but more salient dimension of stimulus variation may acquire exclusive control over behaviour (Lea et al. 2009). Accordingly, the experiments were designed to enable us to dissociate the effects of cue salience from those of cue validity.

Von Fersen and Lea (1990) noted that categorization in pigeons can be investigated in two different ways: by using natural photographs, videos and objects, which may be referred to as an analytic approach, and by using artificially designed stimuli, the synthetic approach. The advantage of using artificial stimuli is that structure and feature content can be carefully controlled, but such control is usually at the expense of the richness and detail associated with natural exemplars. The use of photographs or videos enhances the verisimilitude of categorization experiments, because it presumably mimics more closely the kind of discrimination that birds have to make in the wild (though it cannot do so exactly, and it cannot necessarily be assumed that the birds recognize the pictures as representations of the corresponding real objects). In the present paper, we used both techniques, so as to gain the advantages of both, and also to provide systematic replication of our main manipulation. Experiment 1 used

natural photographs (though with some artificial modification), while Experiment 2 used wholly artificial stimuli In both experiments, the stimuli included a single feature that was a perfectly reliable predictor of category membership; by analogy with the work of Waxman and others cited above, we refer to this as the "label". In addition the stimuli included several other features that were individually imperfect predictors but in combination provided enough information to allow perfect discrimination. Once good performance had been achieved, probe trials were introduced, in which the single, perfectly reliable cue and the remaining cues were put in conflict with each other, allowing us to discern which cues were controlling behaviour more strongly.

Experiment 1

The first experiment used photographs of everyday objects as stimuli. The two categories were houses and cars, which differ along multiple stimulus dimensions. To provide a single highly salient "label" cue, the roofs of the houses and the windows of the cars were replaced with a standard uniform texture in a contrasting colour. If the pigeons in this situation base their discrimination on the label alone, then when probe stimuli are introduced in which the labels have been reversed between categories, it is expected that probes containing the label previously associated with the positively reinforced category (i.e. positive texture probes) will be treated as positive. If pigeons respond positively instead to probes which contain the label originally associated with the negatively reinforced category (positive picture probes), this would suggest they are attending more to the overall similarity of exemplars. Although the stimuli were based on natural photographs of objects, it was not required by the design that the

pigeons should recognize them as depictions of objects that they had seen; photographs were used only to ensure that the stimulus sets had the kinds of variation that are characteristic of functional categories in the everyday world.

Materials and Methods

Subjects

Twenty pigeons were used. They were acquired as discards from pigeon fanciers, and so had visual experience of the world outside the laboratory. Some had previous experience of the touchscreen arrangement used in the present experiment, but in experiments with monochrome, geometric stimuli and a different training procedure (Wills et al. 2009, Experiment 1a); the remainder were experimentally naive. The pigeons were housed in an indoor aviary, measuring 2.2 m by 3.4 m by 2.4 m, and given constant access to water and grit. Prior to testing, pigeons were held in individual cages in which they had access to water and were released back into the aviary when testing had finished for the day. All pigeons were maintained on a 12:12 hr light/dark cycle at 95% of their free-feeding weight.

Apparatus

Four identical operant chambers (internal size; 640 mm x 430 mm x 470 mm) were used. Each consisted of a plywood box, with a 15-inch (39-cm) touch-monitor (Elo Touchsystems Accutouch (resistive) or CarrollTouch (infra-red) model 1547L) mounted in the front wall. The screen resolution was set to 1024 x 768 pixels. The monitor was controlled by software written in Visual Basic using the Whisker control

system (Cardinal and Aitken 2001) running on a computer supplied by Quadvision (Quadvision Ltd., Dorset, UK). Two food hoppers, positioned one on each side of the main screen, could be used to deliver a 2:1 mixture of hemp seed and conditioner to the pigeons, for 2.5 seconds. Each box had a webcam fitted into the side wall, 250 mm above the floor, allowing the pigeons' behaviour to be observed from outside the test room using the imaging software ViewCommander (Internet Video and Imaging, Ltd.). Each pigeon was assigned its own test chamber for all stages of the experiment: six pigeons used resistive touchscreens while the remaining pigeons used infra-red touchscreens.

Stimulus Materials

The images used were natural photographs comprising two stimulus sets; houses and cars. There were 24 pictures of each type of object. Photographs were manipulated using the GNU Image Manipulation Program (GIMP ver. 2.2). Each image was isolated from the original setting, placed on a black background and then scaled to ensure images were approximately the same size (160 x 120 pixels). To produce a highly salient "label" feature, similar areas in the upper part of the objects were selected and given a new artificial texture. For houses this was the roof, and for cars the windows. Thus 12 of the houses had their roofs replaced with a leopard-skin texture, and the other 12 had their roofs replaced by a blue "swimming pool" wave texture. Similarly, 12 of the cars had all their windows replaced with the leopard-skin texture and the other 12 had them replaced with the blue wave texture. Figure 1 shows examples of the stimuli. Fourteen of the pigeons were trained to discriminate the houses with leopard-skin roofs from the cars with blue windows, and for these pigeons probe stimuli consisted of houses with blue roofs and cars with leopard-skin

windows. For the remaining pigeons, the training stimuli were the houses with blue roofs and the cars with leopard-skin windows, and the probe stimuli were the houses with leopard-skin roofs and the cars with blue windows. Within each of these groups, some of the pigeons were trained with houses as positive stimuli and the remainder with cars as positive stimuli.

General Procedure

The pigeons were pre-trained using conventional methods to find food in the grain-feeders, and to peck lighted discs on the touch-screen for food reinforcement. They were then trained in a multi-stimulus discrimination procedure similar to that described by Huber et al. (2005) and Wills et al (2009, Experiment 2b). Initially, they were trained using this procedure to discriminate white filled hexagons (8mm across) within a 25-mm diameter black circle, outlined in white, from a blank circle. Those pigeons that mastered this discrimination proceeded to the task involving the house and car stimuli.

In the multi-stimulus discrimination procedure, each session started when the house-lights in the box came on, and a white disc of diameter 4cm (the observing key) was displayed on the screen. When the pigeon pecked the observing key it disappeared, to be replaced by an array of photographs. The array consisted of eight cells, arranged as shown in Figure 2. In order to 'open' a cell, pigeons had to peck it twice in rapid succession, causing the image to disappear. If the opened cell was positive, a side key appeared to one side of the array, in the nearer of the two locations shown in Fig. 2 (arbitrarily, the left side key was used when the middle top cell was opened, and the right side key when the middle bottom cell was). The pigeons were required to peck

the side key to activate the food hopper on that side for 2.5 seconds (Figure 2). If a negative cell was pecked this was recorded as a miss; the image disappeared but no side key appeared and the screen became unresponsive for 2.5 seconds. If a pigeon pecked a blank cell this was recorded but there were no scheduled consequences. When all the positive cells had been opened, the array disappeared, to be replaced after an inter-trial interval that varied from 1 to 5 s by the observing key. Sessions consisted of six or seven arrays depending on the stage of the experiment.

Training

Three pigeons failed to complete pre-training and were dropped from the experiment. Of the remaining 17 pigeons, 11 were trained using the houses with leopard-skin roofs and the cars with blue windows; of these, five were assigned houses as the positive stimulus, and six were car-positive. The remaining six pigeons were trained using the houses with blue roofs and the cars with leopard-skin windows, and of these three were house-positive and three were car-positive. Arrays were made up of four cars and four houses, pseudo-randomly arranged, with the constraints that stimuli from the same set were never presented in the same place in more than three consecutive arrays and that no more than three stimuli from the same set were placed next to each other or reinforced on the same side. Each training session contained six arrays, so that each of the 12 positive and negative stimuli was seen twice within a session. Discrimination during the training phase was determined using the ρ statistic of Herrnstein et al. (1976), which was used to measure the overlap between stimulus sets in terms of the order in which the cells were opened. A ρ value of 0.5 indicates random responding, whereas a value of 1 indicates perfect discrimination (i.e. that all the positive stimuli have been opened before any of the negative stimuli). After each

pigeon had reached a criterion of a ρ value of 0.8 or more in all six arrays in a session, it was exposed to probe sessions.

Probe Sessions

Probe sessions consisted of four training arrays, alternated with three probe arrays. Each session began and ended with a training array. A probe array contained four training stimuli, two from each stimulus set, and four probe stimuli in which the artificial manipulations were reversed. Two of the probes in each array, one from each stimulus set, were assigned positive reinforcement contingencies and the other two were assigned negative contingencies. Probes that were designated positive the first time they were displayed were not reinforced when next shown. Pigeons were required to open the two positive training cells and all four probe cells in order to complete a probe array. Each pigeon received two probe sessions, which were alternated with training sessions, to ensure pigeons recovered their original training performance. Following the procedure of Wills et al. (2009, Experiment 2b), responses to probe stimuli were categorised according to the order in which the pigeons opened them within the array. A probe that was one of the first four stimuli opened in an array was considered to have been treated as positive, and a probe that was one of the last four stimuli opened was considered to have been treated as negative.

Results

The pigeons reached criterion in a median of 5 training sessions (range 2-8 sessions). This corresponds to the presentation of a median of 30 arrays, i.e. 240 stimuli. There were no significant differences in the rate of acquisition between the house-positive and the car-positive groups, or the leopard-skin positive and blue-wave positive groups, nor was the interaction between the positive picture and the positive texture significant.

In probe sessions, responses to training stimuli continued at the same high level of accuracy. Different pigeons behaved differently towards the probe stimuli. Probes included either a positive picture (house or car, depending on which was positive during training) and a negative texture (blue wave or leopard-skin), or a negative picture and a positive texture. Figure 3 shows, for each pigeon, the proportion of probe stimuli that were treated in accordance with the texture cue present (that is, responded to as positive if they included the positive texture or as negative if they included the negative texture). In all, 7 of the 17 pigeons (Ro, Ma, Fr, Sn, Su, Ba and Ze) responded to more than half the probes in accordance with the texture cue, 9 pigeons (Io, Le, Ri, Ru, Rs, Ti, Ot, Ry, and Se) responded to more than half the probes in accordance with the picture cue, and the remaining pigeon (At) treated both kinds of probes equally. Chi-square tests were used to determine whether the proportions of probe stimuli treated in accordance with the texture or picture cue differed significantly from 0.5 for individual pigeons. Four pigeons responded to the probes in accordance with the texture cue on significantly more than half the trials (Ro, Fr, Sn and Su; χ^2 values of 15.04, 14.09, 8.52 and 4.34, df=1, $p<0.0001$, 0.0001, 0.01 and 0.05 respectively). Three pigeons responded to the probes in accordance with the picture cue on significantly more than half the trials (Io, Ri, and Ry; χ^2 values

of 15.04, 9.38 and 7.04, df=1, p<0.0001, 0.01 and 0.01 respectively). The sum of the individual χ^2 values across the pigeons was 83.05; by virtue of the additive property of the chi-square distribution (Weatherburn, 1957, p. 177), this can be tested as a χ^2 value against the sum of their degrees of freedom. The resulting significance level is < 0.0001, showing that despite the fact that not all results were in the same direction, we can reject the null hypothesis that all pigeons had a 50% chance of treating any probe in accordance with the picture cue.

Discussion

The rapid learning shown by all the pigeons confirms the conclusion of Huber et al. (2005) and Wills et al (2009) that multi-stimulus training methods are an efficient means of establishing complex pattern discriminations. It is a property of such methods, as they have been implemented previously and in the present experiment that, within an array, a correct response makes the remainder of the task more difficult (because it reduces the ratio of positive to negative stimuli remaining) and an incorrect response makes it easier. This makes the task of predicting the probability of a correct response mathematically more difficult, but is not a cocnern when as here, the intention is to compare the probabilities of choosing different probe stimuli that are presented at the same frequency.

Responses to probe stimuli were reinforced non-differentially. This was because substantial numbers of probe trials were required, and the alternative, non-

reinforcement, could have led to the pigeons learning that all probe stimuli were essentially negative. Non-differential reinforcement carries the alternative risk that subjects will come to behave at random towards probes. However in the present experiment, sustained non-random behaviour towards probes was observed in most pigeons, showing that non-differential reinforcement had no major effect.

The pigeons were almost equally split between showing greater control by the picture and showing greater control by the texture (the "label" cue). This was not a result of random behaviour, but of systematic behaviour that differed between individuals. Seven of the 17 pigeons showed significant differences in their responses to probes containing the positive picture rather than the positive texture, but in three of these it was the picture cue that controlled probe responding, and in the other four it was the texture cue. As regards the remaining pigeons, given that performance on training stimuli remained highly accurate throughout the probe sessions, the most plausible conclusion is that they were under the control of both the picture and the texture. The distribution of control by the picture or texture was not even between the groups, and in particular all four of the pigeons that showed a significant trend to texture control had the leopard-skin texture as positive. If the leopard-skin was more salient than the blue-wave texture for the pigeons, this could be accounted for as a feature-positive effect (Jenkins and Sainsbury 1970), but in the absence of independent evidence about relative salience, this can only be a speculation. Furthermore one of the pigeons that showed a significant trend towards control by the picture cue (Io) also had the leopard-skin texture as positive, weakening the argument that the overall pattern of results could be due to high salience of the leopard-skin cue.

Because the pictures were natural photographs, we cannot easily specify which features were supporting discrimination between them, but inevitably the features available will have varied somewhat between instances. We therefore interpret discrimination based on the picture cue as categorization by overall similarity or family resemblance. The texture cue, by contrast, was (to the human eye at least) highly salient, and it was consistent across the training and probe stimulus sets. Discrimination of the texture cue can therefore be regarded as unidimensional categorization. Nevertheless, it only exerted dominant control over behaviour in probe trials for four pigeons.

It is possible that the pigeons learned to discriminate the categories by learning each exemplar of a house or car separately. Previous results make this unlikely given the number of exemplars used: an experiment by Bhatt, cited by Wasserman and Bhatt 1992, showed that pigeons switch from exemplar control to featural control in category discriminations where the number of exemplars rises above about six. In any case it would not matter to the present experiment, whose point was to investigate whether the pigeons would come under the control of the single feature provided by the texture cue or the multiple features provided by the picture cue: if the pigeons discriminated houses from cars on the basis of exemplars, that would involve the learning of even more different features than doing so by extracting a few general features that were positively but not perfectly correlated with reinforcement.

The roughly equal distribution of subjects between control by the texture cue and control by the picture cue is an unexpected result, given that pigeons have a tendency to be dominated by single features (Lea et al. 1993, 2006). However, while it is true

that birds rarely use all the features available, they can certainly be trained to use more than one (e.g. von Fersen and Lea 1990; Lea et al. 2006). Indeed, some of the pigeons used in the present experiment had been trained in Experiment 1a of Wills et al. (2009), in which they had to learn to discriminate several different features presented on separate trials (though the stimuli and the procedure were very different from those of the present experiment). In the generalisation tests of that experiment, when the features were combined, the behaviour of one of the pigeons that was re-used here (Io) was shown to be under the control of multiple features. However, bringing behaviour under the control of multiple stimulus dimensions often requires special training procedures (as in von Fersen and Lea 1990) or extended training, whereas the pigeons in the present experiment learned very quickly.

Because the picture cues were based on natural photographs, it remains an assumption that discrimination of the pictures was on the basis of overall similarity. It is possible that within the pictures there were other highly predictive features apart from the one manipulated, which were consistent within each category and which were salient for the pigeons. All the car stimuli, for example, contain wheels and so pigeons might have learnt the discrimination on the basis of the presence or absence of silver ellipses. We did our best to ensure there was no such single predictive feature for each category, for example the angle at which the photographs of the stimuli were taken was varied, as was the colour of the cars (so that although all the cars had wheels, those wheels themselves formed a polymorphous set), but there can be no certainty that such attempts will succeed. Conversely, although the label features were much more consistent, they did have some variability: for example, the shapes of the house roofs and car windows to which the textures were applied varied between

instances. To avoid these problems while investigating the question of whether birds are more likely to rely on single features or family resemblance, in Experiment 2 we took the alternative approach of using artificial compound stimuli, and so bringing feature content under control.

Experiment 2

Experiment 2 was formally very similar to Experiment 1, but used artificial multidimensional stimuli. The stimuli in the present experiment were made up of four spatially separated elements, each of which constituted a stimulus dimension; they were based on those used by Wills et al. (2009, Experiment 1a) and Lea et al. (2009). The elements are referred to as the Annulus, the Bar, the Checks and the Diamond, and examples are shown in Figure 4. Each element could be used in four graded versions, two of them positively and two negatively correlated with the occurrence of reinforcement. The stimulus sets had exactly the same formal structure as those used by Kemler Nelson (1984) with children and Couchman et al (2010) with rhesus monkeys, but the stimulus dimensions used were different.

Corresponding to the reliable, salient, texture cue used as a label in Experiment 1, one of the dimensions was designated as a label cue for each pigeon. This dimension was used only in its extreme versions, and one of these always occurred in the presence of reinforcement while the other one never did. The remaining three dimensions were used to construct a two-out-of-three artificial polymorphous stimulus of the sort introduced by Dennis, Hampton and Lea (1973) and used with pigeons by Lea and

Harrison (1978): that is, whenever at least two of the three dimensions took its positive value, the stimulus as a whole was designated as positive. For these dimensions, all four versions were used. These three dimensions collectively corresponded to the picture feature in Experiment 1: if a pigeon was to discriminate on the basis of these cues, it would have to come under the control of the overall similarity of the stimuli to an ideal positive form if it was to achieve 100% correct performance, since each individual dimension within this set was imperfectly correlated with reinforcement. These artificial stimulus sets had several additional advantages. First, the spatial separation of the stimulus elements ensured that the dimensions of the stimuli could be manipulated entirely independently. Second, the four dimensions could be used in a balanced way, with each dimension being used to provide the label (reliable) feature for some pigeons. Thirdly, within the set of features used to create the polymorphous concept, it was possible to assess empirically whether all three features were controlling behaviour, and therefore whether it is appropriate to describe the pigeons as coming under the control of overall similarity. Finally, on the basis of results with similar stimuli (Lea et al 2009, Wills et al 2009) we had reason to think that the saliences of the feature differences used on the four dimensions of the stimuli were comparable.

Materials and Methods

Subjects

Sixteen pigeons were used in this experiment. None of them had previous experience of this kind of discrimination task. They were maintained under the same conditions as the pigeons used in Experiment 1.

Apparatus

The same apparatus was used as in Experiment 1, except that only infra-red touchscreens were used.

Stimuli and Experimental design

The stimuli were modified from those used in Wills et al. (2009, Experiment 1a) and Lea et al. (2009). Examples are shown in Figure 4. All stimuli consisted of a square array of four elements. Elements of the same type were always placed in the same location. The element types were an annulus (A), a horizontal bar (B), a square chequerboard (C), and a diamond shape made up of equal-width stripes (D). There were four versions of each element, designated as the X, x, y, and Y forms. All versions of all elements were placed on a black background, and contained the same number of white pixels (within 2%). The specifications of the four forms of each element are listed in Table 1, and they are illustrated in Figure 4. Note that because of the constraint that all elements should have the same area, some elements varied on two dimensions simultaneously: for example, when the Annulus was made smaller, it was also made wider. With four versions of each of four elements, there were 256 possible stimuli, but not all pigeons experienced all stimuli. All versions of all elements fitted within a 60 x 60 pixel square with some black border, so that the entire stimulus including borders fitted within a 120 x 120 pixel square; at a pigeon's typical pre-peck viewing distance, 120 pixels subtended about 25° of arc, and would thus fit within the extent of the pigeon's frontal, binocular viewing field (Martin and Young, 1983).

For each pigeon, one of the four dimensions was designated as Reliable, and either the X or the Y value of it was designated as positive; the less extreme (x and y) values of the Reliable dimension were not used either in training or in probe stimuli. The remaining three dimensions were designated as Unreliable, and either all their X and x values, or all their Y and y values, were designated as positive. Positive training stimuli always included the extreme positive value of the Reliable dimension, and positive values (either extreme or less extreme) of at least two of the Unreliable dimensions. Negative training stimuli always included the extreme negative value of the Reliable dimension, and negative values (either extreme or less extreme) of at least two of the Unreliable dimensions. As an example, consider Pigeon Mo, for which the Reliable dimension was the Annulus and the X values of both the Reliable and Unreliable dimensions were designated as positive (see Table 2). For this pigeon, positive stimuli always contained the X value of the Annulus, and at least two of the other three dimensions (Bar, Chequerboard and Diamond) in either their X or their x forms. The negative stimuli always contained the Y form of the Annulus, and either the y or the Y form of at least two of the other three dimensions. There were 32 stimuli in each of the positive and negative categories. These categories could be discriminated perfectly in either of two ways (or by a mixture of them). The pigeon could use the Reliable dimension alone, and ignore the three Unreliable dimensions. Alternatively, it could ignore the Reliable dimension, and respond on the basis of the preponderance of values of the three Unreliable dimensions. Each Unreliable dimension took one of its negative values in a quarter of the positive training stimuli, and a positive value in a quarter of the negative training stimuli, so each Unreliable dimension considered on its own can be described as being 75% valid during training; their collectivity, however, was 100% valid. The Reliable cue thus served as a label,

while the Unreliable cues defined a 2-out-of-3 polymorphous stimulus set. Stimuli that included the positive value of the Reliable dimension with negative values of at least two of the Unreliable dimensions, or the negative value of the Reliable dimensions with positive values of at least two of the Unreliable dimensions, were ambiguous, and were not used in training. There were 64 such ambiguous stimuli for each pigeon, and a selection of these were used as probes, in tests conducted after training was complete.

The way the categories were used was varied between pigeons so as to balance the use of the different dimensions, as shown in Table 2. Each dimension was assigned as Reliable for four of the pigeons. For two of these, the X value of the Reliable dimension was assigned as positive, and for the other two its Y value was assigned as positive. For one of the pigeons for which each value of the Reliable dimension was assigned as positive, the X and x values of the Unreliable dimensions were assigned as positive, and for the other one the Y and y values of the Unreliable dimensions were assigned as positive.

To reduce the risk that the pigeons would learn the contingencies associated with probe stimuli, only 36 probe trials were given to each pigeon, so not all the 64 possible probe stimuli were used. The stimuli used as probes are summarised in Table 3. Four of these stimuli were used 6 times each, so as to provide a strong test of the basic question of whether the pigeons' behaviour was governed by the Reliable or the Unreliable dimensions. These stimuli involved:

The positive value of the Reliable dimension, combined with the extreme negative values of all three Unreliable dimensions

The negative value of the Reliable dimension, combined with the extreme positive values of all three Unreliable dimensions

The positive value of the Reliable dimension, combined with the less extreme negative values of all three Unreliable dimensions

The negative value of the Reliable dimension, combined with the less extreme positive values of all three Unreliable dimensions.

In addition six other probe stimuli, as indicated in Table 3, were used twice each. These stimuli involved the positive value of the Reliable dimension and the extreme positive value of one of the Unreliable dimensions, and the extreme negative values of the other two Unreliable dimensions; or the negative value of the Reliable dimension and the extreme negative value of one of the Unreliable dimensions, and the extreme positive values of the other two Unreliable dimensions. Each of the Unreliable dimensions was paired with the Reliable dimension in an equal number of probe stimuli. These stimuli allowed a test of which of the Unreliable dimensions was controlling behaviour. Each probe stimulus was associated an equal number of times with the reinforcement contingencies appropriate to positive and negative stimuli.

Procedure

The pigeons were pretrained as in Experiment 1. They were then trained, using the same procedure as in Experiment 1, to discriminate between two positive and negative training categories. Training was continued for a maximum of 20 sessions, but was stopped earlier if a pigeon reached a criterion of a ρ value of 0.8 or more in all six arrays in a session. Three probe sessions where then given. As in Experiment 1, probe sessions consisted of seven arrays: four training arrays, alternated with three

probe arrays. Each session began and ended with a training array. A probe array contained four training stimuli and four probe stimuli, of which two were associated with the reinforcement contingencies appropriate to positive stimuli and two with those appropriate to negative stimuli. Responses to probe stimuli were classified as positive if they occurred within the first four stimulus cells pecked in an array, and as negative otherwise.

Results

Ten of the 16 pigeons reached criterion before their 20^{th} training session; the number of sessions required ranged from 9 to 18 (median, including the pigeons that did not reach criterion, 17). The other six pigeons were showing no obvious further improvement in performance after 20 sessions, though all but one of them was showing ρ values consistently above 0.5; the performance of the remaining pigeon was erratic.

For comparison with Fig. 3, Fig. 5 shows the proportions of probe trials in which the response was correctly predicted by the Reliable stimulus dimension rather than the preponderance of the Unreliable dimensions. For 10 of the 16 pigeons, this proportion was greater than 0.5, and over the group as a whole the deviation from 0.5 was significant (1-sample, 2-tailed Wilcoxon test, $T = 23.5$, $N=16$, $p<0.05$). For each of these pigeons individually, the proportion deviated from 0.5 significantly (2-tailed binomial test). For one of the six pigeons where the proportion was below 0.5 (Ba), the deviation was significant (2-tailed binomial test, $p < 0.001$). The pigeons that did not respond according to the Reliable stimulus on the probe trials had taken

587 significantly longer to reach criterion than those that did (2-tailed Mann-Whitney test,
588 $U = 10.5$, $N_1 = 6$, $N_2 = 10$, $p = 0.03$).
589
590 Figure 6 shows, for each pigeon, the proportions of probe trials in which the response
591 was correctly predicted by each dimension of the stimulus, whether it was reliable or
592 unreliable for that pigeon. In this figure, any proportion above 0.67 or below 0.33
593 would be significantly different from 0.5 on an individual test, though the number of
594 data points involved and the correlations between dimension values mean that simple
595 significance tests cannot be interpreted directly. It can be seen that for each of the
596 pigeons where the unreliable dimensions predicted the response to probe trials
597 markedly better than the reliable dimension (Mo, Bn and Jk), there was one of the
598 unreliable dimensions that predicted response particularly well (Checks for Mo and
599 Jk, and Annulus for Bn). The same trend can be seen more weakly in two of the
600 pigeons where the dominance of the unreliable dimensions was more marginal, Mr
601 and Sa, where Checks and Annulus respectively seemed to have majority control over
602 behaviour. The remaining pigeon, Cr, showed apparently random behaviour towards
603 probe stimuli. For the pigeons where behaviour towards the probe stimuli was
604 dominated but not 100% predicted by the reliable dimension, there was no evidence
605 that individual unreliable dimensions contributed disproportionately to controlling
606 behaviour.
607
608 Because the values of the unreliable dimensions used in the training stimuli varied, it
609 was possible to examine the relative control over behaviour of these dimensions under
610 training as well as probe conditions. Table 4 shows the results of such an analysis,
611 carried out on the training stimuli that were used within the probe sessions so as to

maximise comparability with the probe stimulus results shown in Figures 5 and 6. There were 96 such trials, so any proportion greater than 0.58 or less than 0.42 would be significantly different from 0.5 in a single analysis, though the same cautions about the number of tests and the non-independence of the dimensions must be applied as with Figure 6. These data confirm the dominance of the Annulus dimension for pigeons Sa and Ba, and weakly support the dominance of Checks for Mr, but do not support the dominance of Checks for Mo or Jk. On the other hand they suggest that Cr may have been somewhat under the control of the Bar dimension. It is notable (and logically inevitable) that in those pigeons where one of the unreliable dimensions exerted disproportionate control over behaviour, overall discrimination performance was poorer than in the other pigeons.

Discussion

In Experiment 1, we found that the texture and picture cues (which we interpret respectively as a single reliable dimension, and a collectivity of presumably unreliable cues) were about equally likely to control behaviour, and in many cases individual pigeons probably came under the control of both types of cue. However the two kinds of cues differed in nature. In Experiment 2, where the same cues were used as reliable and unreliable dimensions in a fully balanced way, we found a clear trend for reliable dimensions rather than a collectivity of unreliable dimensions to control behaviour, even though either of them could have enabled the pigeons to perform with perfect accuracy. There were some individuals in which the unreliable dimensions exerted dominant control over behaviour, but in at least some cases it appears that this was not because the collectivity of those dimensions was enabling perfect discrimination, but rather because one of those dimensions was controlling behaviour

637 to the exclusion of the others and of the reliable dimension, with the result that the
638 pigeon's discrimination accuracy was limited. The pigeons that did not come under
639 the control of the Reliable dimension were slower to reach criterion (indeed, 4 out of
640 these 6 pigeons did not reach it within 20 sessions, as against only 2 of the 10 pigeons
641 that did show Reliable dimension control), and if they were coming under the
642 exclusive control of a less valid dimension, this is a more or less inevitable result.
643 Our results thus differ from those obtained by Couchman et al (2010), using stimulus
644 sets that were structurally identical to those used here, though with different elements:
645 Compared with humans exposed to the same task, Couchman et al's rhesus monkeys
646 always showed a greater tendency to categorise novel stimuli in terms of overall
647 similarity rather than in terms of the label dimension, and the authors concluded that
648 this was because the monkeys could not use verbal rules to categorize complex
649 stimuli. Since we do not believe that pigeons use verbal rules, yet in our experiment
650 they showed a clear tendency to use the label dimension rather than family
651 resemblance, we conclude that categorization by family resemblance is not an
652 inevitable consequence of failing to use verbal rules. We therefore also conclude that
653 unidimensional categorization is not a reliable sign that a verbal rule is being used, in
654 accordance with the conclusions of Lea and Wills (2008) and Wills et al (2009). The
655 difference between our results and those of Couchman et al may be due to the subject
656 species, or it may be due to differences in the details of the stimuli involved, a matter
657 to which we return in the General Discussion.
658
659 Part of the reason why clear dominance of the reliable dimension was found in the
660 present experiment is that with the values on them used in the present experiment, the
661 salience of the four dimensions seems to have been, if not equal, at least comparable,

as we predicted on the basis of our previous results with similar stimuli (Wills et al 2009, Lea et al 2009). As Table 4 shows, when used as the reliable dimension, all four dimensions achieved dominant control over behaviour in at least one of the four pigeons for which they were Reliable; assuming that salience is reflected in the number of pigeons for which this was so, the order for salience was Checks > Annulus> Diamond > Bar. The data on dominance by Unreliable dimensions (also shown in Table 4) confirm this pattern, with Checks and Annulus being the only dimensions to achieve dominant control when unreliable. Acquisition data also show that the pigeons for which these dimensions were Reliable tended to learn more quickly than the others.

Learning in Experiment 2 was substantially slower than in Experiment 1. In Experiment 1, pigeons took a median of 5 training sessions (range 2-8 sessions) to reach criterion; in Experiment 2, the median number of sessions to the same criterion was 12, with the number required ranging from 3 to over 20. This difference is consistent with the fact that in several studies where polymorphous concept acquisition has been slow or incomplete artificial stimuli have been used (e.g. Lea et al., 1993, 2006), whereas otherwise quite similar studies using natural photographic stimuli have found more rapid acquisition (e.g. von Fersen and Lea, 1990). It may be that there is something about abstract geometric stimuli that makes it hard for pigeons to learn about them. The present results do however rule out one explanation for the ease of learning discriminations based on natural photographs, which is the possibility that natural stimuli contain artefacts that enable the subjects to make an easy single-dimensional discrimination, whereas the more fully controlled artificial stimuli do not.

In Experiment 2, all the discriminations could have been solved by the use of a single dimension of demonstrated salience, and yet they were not easy for the pigeons.

General Discussion

Both of the present experiments examined the effect on pigeons' category discrimination of providing a single, salient dimension, in effect a category label, alongside more variable information that was nonetheless sufficient when taken as a whole to support perfect discrimination. In Experiment 1, the label was a distinctive texture, applied to parts of natural photographs. In Experiment 2, it was one of four artificial geometric elements, distinguished by the fact that it had 100% validity as a cue to reward, whereas the other elements only had 75% validity individually, though collectively they were fully valid. In Experiment 1, some pigeons clearly based their discrimination on the label, but slightly more of the pigeons used the pictorial information instead. In Experiment 2, on the other hand, almost all the pigeons used the label (the Reliable feature); of the small number that did not, most came under the control of just one of the Unreliable features, and were thereby prevented from achieving accurate discrimination.

The results of Experiment 1 demonstrate a tendency that is evident in the literature on complex visual discriminations in pigeons; that detailed photographic material can serve as highly effective stimuli in such discriminations, competing in effectiveness with simple unidimensional cues on highly salient dimensions such as colour. On the other hand, taking the results of the two experiments together, they do not support the idea that pigeons discriminate photographs readily because the birds have a strong tendency to integrate the wealth of different and possibly unreliable cues that pictures

711 contain. Rather, the results support the conclusions of Lea and Wills (2008), Wills et
712 al. (2009) and Lea et al. (2009), that pigeons have a strong tendency to come under
713 the control of single salient cues when exposed to complex discrimination tasks.
714 What photographs offer is a wide range of strong cues, so there is a good chance that
715 any pigeon will find one that is salient for it – salience being presumably, in part, a
716 function of the individual's previous experience. The carefully balanced design used
717 in Experiment 2 enables us to conclude that, other things being equal, if there is a
718 fully reliable cue available it is likely that it will achieve dominant control over
719 behaviour. Nonetheless, it is not certain that this will happen; it is possible for more
720 salient but less reliable cues to dominate, resulting in imperfect discrimination. The
721 extreme case of this is the occurrence of position habits in experiments where animals
722 have to make spatial choices in a discrimination task; by the design of the experiment,
723 position is normally a cue that carries no information at all, but it can still achieve
724 dominant control over behaviour that continues despite evidence that an animal has
725 detected the truly predictive contingencies (e.g. Mahut 1954).
726
727 Despite the strong tendency for pigeons to come under the control of single
728 dimensions, it is clearly not the case that they cannot be controlled by more than one
729 dimension. Nor is it the case that control by multiple dimensions only occurs when it
730 is essential for perfect discrimination, as in a compound discrimination. These
731 generalizations were again confirmed in the present experiments: In Experiment 1, it
732 appeared that a majority of the pigeons came under the control of both the picture and
733 the texture, and in Experiment 2 at least a few of the pigeons showed evidence of
734 control by more than one of the unreliable dimensions.
735

While control by multiple dimensions is not impossible, it seems from the present results, and previous data, that it is not the natural tendency for pigeons. It may be that it is easier for primates, and this would be one account of the difference between our results and those of Couchman et al (2010) with rhesus monkeys. If multidimensional control does not come naturally to pigeons, tasks that require it, such as polymorphous discriminations, are likely to be difficult for pigeons or other birds to learn, and previous data support this position (e.g. Lea and Harrison 1978; Lea et al. 2006). Lea et al. concluded that the reason was a limitation of attention: any task that requires the processing of more than one stimulus dimension is inherently difficult for a bird. The present data are consistent with that position.

It is possible that the difficulty of attending to multiple dimensions is exacerbated when the dimensions are spatially separated, as in the present experiments. It is notable that in the stimuli used by Couchman et al (2010), who obtained categorization by overall similarity from rhesus monkeys, the four elements were spatially contiguous, and this could be an alternative account of the difference between their results and ours. Spatial grouping does affect pigeons: for example, Sainsbury (1971) found that the feature positive effect in pigeons, which depends on the elements of a stimulus being processed separately, was attenuated when the elements were grouped more closely together. It is also possible that some kinds of stimulus dimension are more separable than others for cognitive rather than perceptual reasons (Soto and Wasserman 2010). However it is not a foregone conclusion that spatial separation or cognitive compatibility will lead to a greater tendency towards categorization by overall similarity: they could instead act to mitigate confusion

between the dimensions. This is a matter that would repay experimental investigation.

The task used in the present Experiment 2 had many points in common with a standard *m*-out-of-*n* artificial polymorphous discrimination task. The essential difference was that there was an additional, completely reliable stimulus dimension (a label in the terms used by Waxman and Markow, 1995), whereas in a standard polymorphous task, all dimensions are equally unreliable. It is therefore not surprising that the task was learned relatively quickly, especially when the complex and abstract nature of the stimuli is taken into account. The speed of learning may have been partly due to the multiple simultaneous presentation procedure, which is known to produce faster learning than a simple go/no-go task (Huber et al. 2005, Wills et al. 2009), and it would be interesting to try a standard polymorphous discrimination using this procedure. However the present design does raise an intriguing possibility. We normally think of the different dimensions of a stimulus as competing for a subject's limited capacity for attention, and this is the basis of most attention-based theories of discrimination learning, e.g. Sutherland and Mackintosh (1971). Indeed, earlier attention-based theories assumed that animals could only attend to a single stimulus dimension at once (e.g. Krechevsky, 1932). Our results are certainly consistent with the idea that animals' learning of complex stimuli is limited by their attentional capacity. It is conceivable, however, that a highly reliable dimension could act to inform an animal that a task can be learned, and this might heighten attention to other dimensions rather than diminishing it, especially if the reliable dimension was removed once learning had taken place; something of this sort might underlie the "labeling" phenomenon as it occurs in young children (Waxman

and Markow, 1995; Balaban and Waxman, 1997; Plunkett et al., 2008). This possibility gives more potential empirical bite to limited-attention theories, which can seem to add little to the empirical generalization that multiple-dimension discriminations are difficult. It could perhaps be investigated by exploring the mechanisms of attention in more detail by comparing the amount learned about unreliable stimulus dimensions in the presence or absence of more reliable cues.

Acknowledgements

Experiment 1 was carried out by Beth Nicholls in partial fulfilment of the requirements for the MSc degree in Animal Behaviour of the University of Exeter. The experimental work described here was supported by the New and Emerging Science and Technologies activity of the European Community Framework Programme 6, under grant no. 516542, "From Associations to Rules in the Development of Concepts" (FAR). We are grateful to Lisa Leaver, Louise Millar, Andy Wills and John Endler for discussion, and to Denis Mareschal and Caspar Addyman (Birkbeck, University of London) for supplying the photographs used as stimuli in Experiment 1. The experimental work was carried out within the terms of the United Kingdom Animals (Scientific Procedures) Act (1986).

References

Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM (1998) A formal neuropsychological theory of multiple systems in category learning. Psychol Rev 105:442-481

Blough DS (1972) Recognition by the pigeon of stimuli varying in two dimensions. J Exp Anal Behav 18:345-367

Bovet D, Vauclair J (2000) Picture recognition in animals and humans. Behav Brain Res 109:143-165

Cardinal, RN, Aitken, MRF (2001) Whisker (version 2). Cambridge University Technical Services Ltd, Cambridge

Couchman JJ, Coutinho MVC, Smith, JD (2010) Rules and resemblance: Their changing balance in the category learning of humans (*Homo sapiens*) and monkeys (*Macaca mulatta*). J Exp Psychol: Anim Behav Proc 36: 172-183

Dennis I, Hampton JA, Lea S (1973) New problem in concept formation. Nature 243:101-102

Herbranson WT, Fremouw T, Shimp, CP (1999) The randomization procedure in the study of categorization of multidimensional stimuli by pigeons. J Exp Psychol: Anim Behav Proc 25: 113-134

Herrnstein RJ, Loveland DH (1964) Complex visual concept in the pigeon. Science 146: 549-551. doi:10.1126/science.146.3643.549

Herrnstein RJ, Loveland DH, Cable C (1976) Natural concepts in pigeons. J Exp Psychol Anim Behav Proc 2: 285-302. doi:10.1037.0097-7403.2.4.285

Huber L (2001) Visual categorization in pigeons. In: Cook RG (ed) Avian visual cognition. [On-line], Available www.pigeon.psy.tufts.edu/avc/huber/

Huber L, Apfalter W, Steurer M, Prossinger H (2005). A new learning paradigm elicits fast visual discrimination in pigeons. J Exp Psychol Anim Behav Proc 31, 237-246. doi:10.1037/0097-7403.31.2.237

Kemler Nelson, DG (1984) The effect of intention on what concepts are acquired. J Verb Lng Verb Behav 23:734-759

833 Kirkpatrick-Steger K, Wasserman EA (1996) The what and the where of the pigeon's
834 processing of complex visual stimuli. J Exp Psychol Anim Behav Proc 22:60-
835 67. doi:10.1037/0097-7403.22.1.60
836 Krechevsky I (1932) Hypotheses in rats. Psychol Rev 39:516-532
837 Jenkins HM, Sainsbury RS (1970) Discrimination learning with the distinctive feature
838 on positive or negative trials. In: Mostofsky DI (ed) Attention: Contemporary
839 theory and analysis. Appleton-Century-Crofts, NewYork, pp 239-273
840 Lea SEG, Harrison SN (1978) Discrimination of polymorphous stimulus sets by
841 pigeons. Q J Exp Psychol 30:521-537
842 Lea SEG, Lohmann A, Ryan CME (1993) Discrimination of five-dimensional stimuli
843 by pigeons: Limitations of feature analysis. Q J Exp Psychol 46B:19-42.
844 doi:10.1080/14640749308401093
845 Lea SEG, Wills AJ (2008) Use of multiple dimensions in learned discriminations.
846 Comp Cogn Behav Rev 3, 115-133. doi:10.3819/ccbr.2008.30007
847 Lea SEG, Wills AJ, Leaver LA, Ryan CME, Bryant CML, Millar L (2009) A
848 Comparative analysis of the categorization of multidimensional stimuli: II.
849 Strategic information search in humans (*Homo sapiens*) but not in pigeons
850 (*Columba livia*). J Comp Psychol 123:406-420 . doi:10.1037/a0016851
851 Lea SEG, Wills AJ, Ryan CME (2006) Why are artificial polymorphous concepts so
852 hard for birds to learn? Q J Exper Psychol 59:251-267 doi
853 10.1080/02724990544000031
854 Mahut H (1954) The effect of stimulus position on visual discrimination by the rat.
855 Canad J Psychol 8:130-138
856 Makino H, Jitsumori M (2007) Discrimination of artificial categories structured by
857 family resemblances: a comparative study in people (*Homo sapiens*) and

858 pigeons (*Columba livia*). J Comp Psychol 121, 22-33. doi:10.1037/0735-
859 7036.121.1.22
860 Martin GR, Young SR (1983) The retinal binocular field of the pigeon (*Columba*
861 *livia*: English racing homer). Vision Research 23:911-915
862 Pothos EM (2005) The rules versus similarity distinction. Behav Brain Sci 28:1-49
863 Sainsbury RS (1971) Effect of proximity of elements on the feature-positive effect. J
864 Exp Anal Behav 16:315-325
865 Smith LB (1981) Importance of the overall similarity of objects for adults' and
866 children's classifications. J Exp Psychol Human Percep Perf 7, 811-824.
867 doi:10.1037/0096-1523.7.4.811
868 Smith LB, Kemler DG (1977) Developmental trends in free classification; Evidence
869 for a new conceptualisation of perceptual development. J Exp Child Psychol
870 24, 279-298. doi:10.1016/022-0965(77)90007-8
871 Soto FA, Wasserman EA (2010) Integrality/separability of stimulus dimensions and
872 multidimensional generalization in pigeons. J Exp Psychol Anim Behav Proc
873 36:194-205 doi :10.1037/a0016560
874 Sutherland NS, Mackintosh NJ (1971) Mechanisms of animal discrimination learning.
875 Academic Press, New York
876 Von Fersen L, Lea SEG (1990) Category discrimination by pigeons using five
877 polymorphous features. J Exp Anal Behav 54, 69-84.
878 doi:10.1901/jeab.1990.54-69
879 Wasserman EA, Bhatt RS (1992) Conceptualization of natural and artificial stimuli by
880 pigeons. In: Honig WK, Fetterman JG (eds) Cognitive aspects of stimulus
881 control. Erlbaum, Hillsdale NJ, pp 203-223

882 Weatherburn CE (1957) A first course in mathematical statistics, 2nd Edn corrected.
883 Cambridge University Press, Cambridge
884 Wills AJ, Lea SEG, Leaver LA, Osthaus B, Ryan CME, Suret MB, Bryant CML,
885 Chapman SJA, Millar L (2009) A comparative analysis of the categorization
886 of multidimensional stimuli: I. Unidimensional classification does not
887 necessarily imply analytic processing; evidence from pigeons (*Columba livia*),
888 squirrels (*Sciurus carolinensis*) and humans (*Homo sapiens*). J Comp Psychol
889 123:391-405. doi:10.1037/a0016216
890

891 Table 1. Experiment 2: The four forms of each dimension of the stimuli

Element	dimension of variation	X value	x value	y value	Y value
Annulus	width:radius ratio	5:1	4:1	2:1	1.5:1
Bar	length:width ratio	7.5:1	6:1	3.3:1	2:1
Chequerboard	number of elements	2 x 2	4 x 4	6 x 6	8 x 8
Diamond	orientation of stripes	vertical	60º	30º	horizontal

892

893

Table 2

Experiment 2: Use of stimulus dimensions in training stimuli for each pigeon

Pigeon	Reliable dimension (label)	Positive value	Unreliable dimensions	Positive values
Mo	Annulus	X	Bar, Chequerboard, Diamond	X, x
Ct	Annulus	X	Bar, Chequerboard, Diamond	Y, y
Ch	Annulus	Y	Bar, Chequerboard, Diamond	X, x
Kc	Annulus	Y	Bar, Chequerboard, Diamond	Y, y
Bn	Bar	X	Annulus, Chequerboard, Diamond	X, x
Mr	Bar	X	Annulus, Chequerboard, Diamond	Y, y
Sa	Bar	Y	Annulus, Chequerboard, Diamond	X, x
Cu	Bar	Y	Annulus, Chequerboard, Diamond	Y, y
Rg	Chequerboard	X	Annulus, Bar, Diamond	X, x
Hy	Chequerboard	X	Annulus, Bar, Diamond	Y, y
Yw	Chequerboard	Y	Annulus, Bar, Diamond	X, x
Bu	Chequerboard	Y	Annulus, Bar, Diamond	Y, y
Fl	Diamond	X	Annulus, Bar, Chequerboard	X, x
Cr	Diamond	X	Annulus, Bar, Chequerboard	Y, y
Dd	Diamond	Y	Annulus, Bar, Chequerboard	X, x
Jk	Diamond	Y	Annulus, Bar, Chequerboard	Y, y

Table 3. Example showing the Probe stimuli used in Experiment 2. In this example, A was the Reliable dimension, with its X value as positive, and B, C and D were the Unreliable dimensions, with their X and x values as positive. The dimensions used as reliable and unreliable, and the values used as positive and negative, varied between birds in a balanced fashion.

Stimulus dimension				
Annulus (Reliable)	Bar	Chequerboard (Unreliable)	Diamond	Number of times used
X	Y	Y	Y	6
Y	X	X	X	6
X	y	y	y	6
Y	x	x	x	6
X	X	Y	Y	2
Y	Y	X	X	2
X	Y	X	Y	2
Y	X	Y	X	2
X	Y	Y	X	2
Y	X	X	Y	2

905 Table 4. Experiment 2: Proportions of training trials within test sessions in which the
906 response was correctly predicted by each dimension of the stimulus

Reliable dimension	Pigeon	Proportions of trials correctly predicted by value of				
		Positive stimulus (Reliable dimension and majority of unreliable dimensions)	Individual unreliable dimensions			
			Annulus	Bar	Checks	Diamond
Annulus	Mo	0.64		0.63	0.58	0.58
Annulus	Ct	0.87		0.76	0.67	0.70
Annulus	Ch	0.98		0.77	0.77	0.78
Annulus	Kc	0.98		0.72	0.79	0.79
Bar	Bn	0.70	0.85		0.55	0.50
Bar	Mr	0.64	0.53		0.60	0.58
Bar	Sa	0.66	0.69		0.58	0.53
Bar	Cu	0.95	0.80		0.68	0.73
Checks	Rg	0.88	0.77	0.61		0.73
Checks	Hy	0.98	0.79	0.73		0.77
Checks	Yw	0.89	0.80	0.70		0.66
Checks	Bu	1.00	0.81	0.76		0.77
Diamond	Fl	0.86	0.72	0.68	0.68	
Diamond	Cr	0.46	0.41	0.61	0.38	
Diamond	Dd	0.86	0.74	0.70	0.68	
Diamond	Jk	0.71	0.77	0.64	0.51	

907

Figure Legends

Fig. 1. Experiment 1: Examples of stimuli showing each of the two picture types (house and car) associated with each of the two artificially introduced textures (leopard-skin and blue wave). Each stimulus measured 160 x 120 pixels. (From original photographs by Casper Addyman, used with permission)

Fig. 2. Diagram of touch screen display, showing size and position of array and two side keys. Cells were numbered 1-8, clockwise from top left. Cells 3-6 were reinforced on the right of the screen: cells 7, 8, 1 and 2 reinforced on the left.

Fig. 3. Proportions of probe stimuli responded to in accordance with the texture cue. Data are shown separately for each pigeon, separated by the stimulus type that was positive in training (*=Difference from 0.5 significant at 0.05 level).

Fig. 4. Examples of the stimulus sets used in Experiment 2. The upper two panels show the A and a versions of each dimension, the lower two panels the b and B versions.

Fig. 5. Experiment 2: Proportions of probe trials in which the stimulus was responded to in accordance with the Reliable stimulus dimension. Data are shown separately for each pigeon, grouped by the dimension that was designated as Reliable (*=Difference from 0.5 significant at 0.05 level).

Fig. 6. Experiment 2: Proportions of probe stimulus trials on which each of the Unreliable dimensions correctly predicted response. Data are shown separately for each pigeon, grouped by the dimension that was designated as Reliable. Gaps within the cluster of bars for each pigeon correspond to the Reliable dimension

Figure 1.

Fig. 1 Experiment 1: Examples of stimuli showing each of the two picture types (house and car) associated with each of the two artificially introduced textures (leopard-skin and blue wave). Each stimulus measured 160 x 120 pixels. (From original photographs by Casper Addyman, used with permission)

a

b

Figure 2

Fig. 2 Diagram of touch screen display, showing size and position of array and two side keys. Cells were numbered 1-8, clockwise from top left. Cells 3-6 were reinforced on the right of the screen: cells 7, 8, 1 and 2 reinforced on the left.

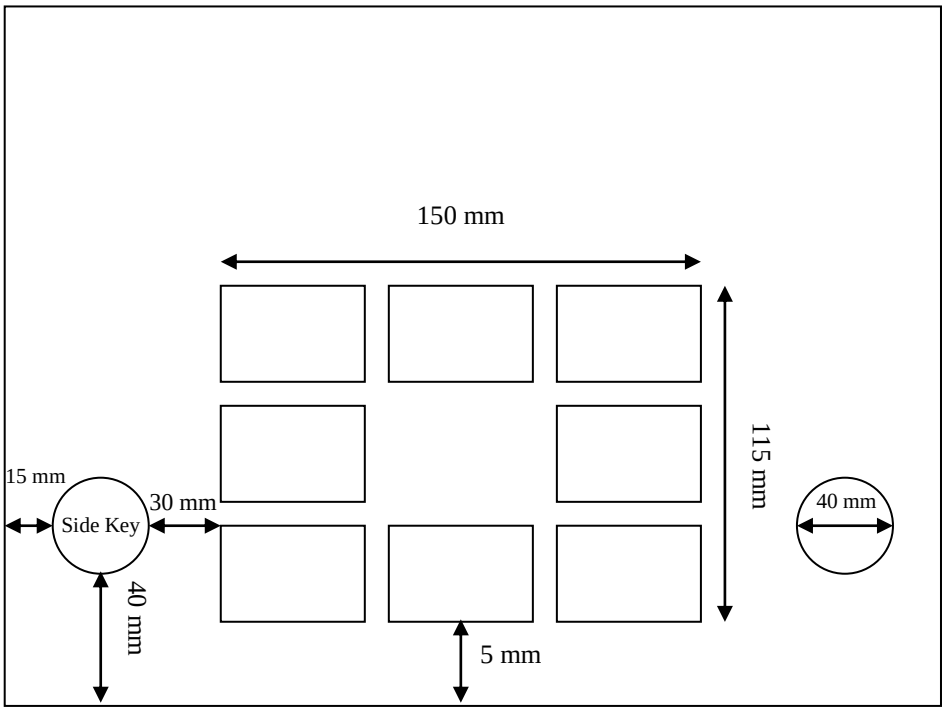

Figure 3

Fig. 3 Experiment 1: Proportions of probe stimuli responded to in accordance with the texture cue by each pigeon, separated by the stimulus type that was positive in training (*=Difference from 0.5 significant at 0.05 level).

[see separate file]

Fig. 4. Examples of the stimulus sets used in Experiment 2. The upper two panels show the X and x versions of each dimension, the lower two panels the y and Y versions.

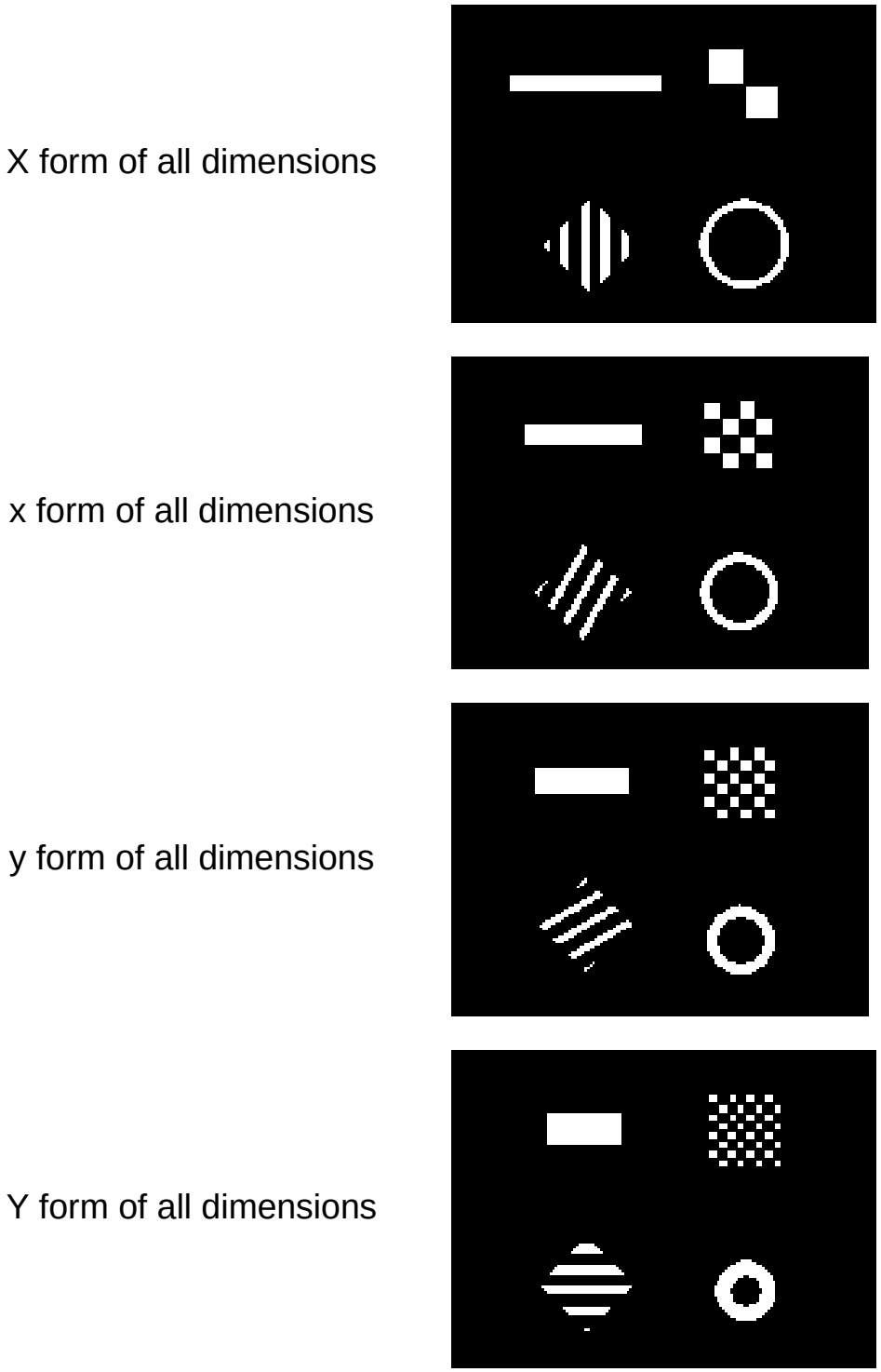

Fig. 5 Experiment 2: Proportions of probe trials in which the stimulus was responded to in accordance with the Reliable stimulus dimension. Data are shown separately for each pigeon, grouped by the dimension that was designated as Reliable (*=Difference from 0.5 significant at 0.05 level). Gaps within the cluster of bars for each pigeon correspond to the Reliable dimension

[see separate file]

Fig. 6. Experiment 2: Proportions of probe stimulus trials on which each of the Unreliable dimensions correctly predicted response. Data are shown separately for each pigeon, grouped by the dimension that was designated as Reliable.